2022 Flash Flood and Intense Rainfall (FFaIR) Experiment: Program Overview and Operations Plan

June 20 - July 22, 2022

Held Virtually

Supported by the Weather Prediction Center

Hydrometeorology Testbed

Table of Contents

1. Introduction

Umbrellaed under the Weather Prediction Center's Hydrometeorology Testbed (WPC-HMT) the Flash Flood and Intense Rainfall (FFaIR) experiment has been held annually since 2012. It works as a foundation to evaluate experimental guidance and tools in a pseudo-operational setting as they pertain to heavy rainfall and flash flood forecasting and is a part of the National Weather Service's (NWS) Research-to-Operations (R2O) activities. Each year it brings together meteorologists and hydrologists across multiple areas of expertise, from forecasters to model developers to academia, to discuss the utility of experimental products and help to foster collaboration and understanding among the various disciplines.

The importance of improving our ability to forecast precipitation was highlighted by NOAA's Precipitation Prediction Grand Challenge, which acknowledges that although many aspects of the forecast have improved, precipitation forecasts have not seen the same improvement. Impacts from precipitation, such as flash flooding, can have far-reaching impacts on both life and property. For instance, according to NOAA's National Centers for Environmental Information (NCEI), of the 20 billion-dollar weather driven disasters, 2 were caused by flooding while another 4 were from tropical cyclones that brought widespread flash flooding to the areas they impacted (ex. Hurricane Ida; NCEI 2021[1]). Through the R2O process, FFaIR and its partners work to help improve the NWS's ability to forecast heavy rainfall and the communication of the risks (i.e. flash flooding) to our partners and the public.

2. Experiment Operations

FFaIR will once again be held virtually and will take place from late June to late July, with a week off for the Fourth of July. New this year, the experiment will also have to accommodate the Juneteenth Federal Holiday, which will be observed on June 20th. The current plan is to move the start date of FFaIR from June 20 to June 21 and have the first week of FFaIR be a short week, with only four days rather than five. Therefore, the four weeks are as follows:

Week 1: June 21 - 24
Week 2: June 27 - July 1
Week 3: July 11 - 15
Week 4: July 18 - 22

For the most part, participants are expected to attend the entire week they are assigned. Along with the daily activities, FFaIR will host a seminar series that focuses on the experimental guidance and tools being evaluated and rainfall/flooding research. These seminars will be open to the entire NWS and various outside partners. The seminars are listed in Table 1.

---

[1] https://www.ncdc.noaa.gov/billions/events/US/2021

Table 1: The 2022 FFaIR seminar schedule.

Dates (2-3 EDT)	Presenter(s)	Topic/Title of Seminar	Affiliation
Tues. - June 7	Jennifer Shoemaker	"Integrating WoFS into May 30th, 2021 Flash Flood Operations"	WFO - ABQ
Tues. - June 14	Jacob Carley	EMC briefing on the development of the RRFS	EMC
Tues. - June 21	Brenda Phillips	Survey results on motorists decision-making and urban flash floods in the DFW area	U Mass and CASA
Thurs. - June 23	Pat Spoden	Forcasting and IDSS for dual weather threats	WFO - PAH
Tues. - June 28	Erik Nielsen and Jen Henderson	Physical and social science aspects of TOR/FF events	Texas A&M and Texas Tech University
Thurs. - June 30	Russ Schumacher and Aaron Hill	New CSU MLP Products - GEFS ERO Trained on the FV3 GEFS and retraining based on a alternative observation dataset	Colorado State University
Tues. - July 12	Marty Baxter	Climatology of Michigan rainfall and evaluation of the July 21, 2019 record rainfall event	Central Michigan University
Thurs. - July 14	Keith Brewster and Tim Supinie	OU CAPS RRFS ensemble performance and discussion of a new machine learning mean prodcut	OU CAPS
Tues. - July 19	Andrew Orrison	A discussion of WPC's METwatch operations	WPC
Thurs. - July 21	Shakira Stackhouse	"Evaluating the Skillfulness of the Hurricane Analysis and Forecast System (HAFS) Forecasts for Tropical Cyclone Precipitation using an Object-Based Methodology"	AFSO-FDS Water Resources Services Branch

2.1 Daily Schedule

Aside from the first day of each week of FFaIR, the experiment will run from 930am to 430pm EDT. By starting later in the morning (relative to eastern time), we hope to be able to better accommodate participants across the western CONUS. The platform that will be used to host virtual FFaIR is Google Meet and its associated tools. The FFaIR team recognizes the stress associated with being on virtual chats for long periods of time. Therefore, aside from the scheduled breaks, the FFaIR team will be encouraging the participants to take a break from the screen whenever they begin to feel strain or stress from screen time.

A general summary of the daily schedule can be seen in Table 2. However, this should be thought of as more of guidelines than a set-in-stone schedule.

Table 2: Tentative daily schedule for the 2022 FFaIR Experiment.

Time (UTC)	Monday	Tuesday	Wednesday	Thursday	Friday
1330 – 1430	Ice Breaker and Orientation	Morning review of yesterday's weather and Day 1 ERO and AERO Forecasting Activity	Morning review of yesterday's weather and Day 1 ERO and AERO Forecasting Activity	Morning review of yesterday's weather and Day 1 ERO and AERO Forecasting Activity	Morning review of yesterday's weather and Day 1 ERO and AERO Forecasting Activity
1400 – 1600	Day 1 ERO and AERO Forecasting Activity				
1600 – 1645	Lunch	Lunch	Lunch	Lunch	Lunch
1645 – 1800	Day 1 MRTP Forecast	Day 1 MRTP Forecast	Day 1 MRTP Forecast	Day 1 MRTP Forecast	Day 1 MRTP Forecast
1800 – 1810	Break		Break		Break
1810 – 1930	Verification		Verification		Verification
1800 – 1900		DIFFERENT GOOGLE LINK - Science Seminar		DIFFERENT GOOGLE LINK - Science Seminar	
1900-2030	Day 2 MRTP Forecast	Verification and Day 2 MRTP Forecast	Day 2 MRTP Forecast	Verification and Day 2 MRTP Forecast	Day 2 MRTP Forecast

Each day will start off with a forecast discussion led by a WPC Forecaster. The discussions include a mix of the previous day's weather, a synoptic overview with operational models and analysis of the differences/similarities among the experimental and operational guidance. Participants will then split into breakout groups to work on their collaborative forecasts - either creating an Excessive Rainfall Outlook (ERO) or an ERO based on Average Recurrence Intervals (AERO). A similar cadance will be followed for the Day 1 and Day 2 Maximum Rainfall and Timing Product (MRTP) forecast activity. The forecaster will lead a discussion relevant to the valid time of the MRTP, then participants will work on their own to create their forecast. It will depend on the flow of the day and the anticipated impact of precipitation on whether or not a Day 2 MRTP is issued. For most days, the forecast will be done in real time. However, because the experimental model/ensemble data being provided to FFaIR rely on each other (this is explained in detail in Section 3), there are instances in which computational issues would result in no experimental data being available. To mitigate the impact this might have on the experiment, the FFaIR team will be collecting cases from the experimental systems when they are run for the Hazardous Weather Testbed (HWT). If needed, days in which the experimental models are not running in real time, participants will forecast using a retrospective case that occurred during HWT (May 2- June 3).

Along with the forecast activities, participants will be tasked with subjectively verifying experimental guidance and tools that are centered around the experimental goals outlined in Section 2.2. This will be done through a variety of ways, including the traditional Quantitative Precipitation Forecast (QPF) verification done by comparing the 24 hr QPF with Multi-Radar

Multi-Sensor Gauge Corrected (hereafter MRMS) Quantitative Precipitation Estimate (QPE). Finally, as already noted, FFaIR will also be hosting a series of seminars that are open to the NWS. Unlike in previous years where the seminars were only held during FFaIR, this year the seminar series will expand beyond FFaIR. The current goal of the FFaIR team is to host the series across all of June and July (minus the week of July 4th). During FFaIR, two seminars will be held each week, on Tuesday and Thursday. During non-FFaIR weeks, the seminars will be hosted on Tuesdays.

2.2 Experiment Goals

The experimental goals provide a general summary of what the FFaIR team plans on focusing the subjective verification questions on. They are by no means the only goals of the experiment and the team hopes that open discussion about the products with the participants will help lead to additional insights about the guidance evaluated that expand beyond the goals listed below. Since our partners will be providing a plethora of Rapid Refresh Forecast System (RRFS) data, a majority of our goals will revolve around this system of models and ensembles.

- Evaluate the usefulness of operational and experimental products from high resolution convective-allowing deterministic and ensemble models' (CAM) QPF. This includes focusing on QPF thresholds exceeding 1inch, as the previous two FFaIR's have noted that at precipitation thresholds greater than 1inch, the wet bias from FV3 convective allowing models increases quickly as the threshold value increases (Trojniak and Correia 2020, 2021[2]).

- Assess the impact of FV3-CAMs configuration changes that may reduce the prolific precipitation and/or precipitation rates related to grid point storms (also referred to as popcorn storms) that were identified in the 2020 FFaIR Experiment and continued to be seen in the 2021 FFaIR Experiment. This will include assessing the hourly maximum precipitation rates out of the FV3-CAMs; only the instantaneous precipitation was accessed last year.

- Evaluation of the impact of cycled data assimilation on the first six hours of the different RRFS members.

- Use the MRTP to identify timing errors during MCS events using the 6 hour precipitation verification. In past FFaIRs, it has been noted that models might have the correct idea of how an event might evolve but do not have the timing correct (ex. initiation, change in progression direction and speed, etc.).

- Analyze the utility of various RRFS ensemble configurations, from multi-physics to stochastic parameter perturbations (SPP), and compare their performances to the HREF. In addition to evaluating "classic" ensemble probabilities, FFaIR will be evaluating a machine learning product for the probability of exceedance from the CAPS group.

[2] 2020 FFaIR Final Report and 2021 FFaIR Final Report

- Evaluation of the Colorado State University's (CSU) Machine Learning Products (MLP) for the Day 1 ERO. This year, two new versions of their operational GEFS-based ERO will be analyzed, as well as an updated version of the HRRR-based ERO from the 2021 FFaIR experiment.
- Evaluate the utility of creating an Excessive Rainfall Outlook that is centered around the exceedance of Average Recurrence Intervals (ARI). These forecasts will attempt to predict locally heavy rain that the traditional ERO may not consider due to small spatial or temporal scales. It will be referred to as AERO and created by the participants.
- In the 2020 FFaIR experiment, a product tracking heavy precipitation objects (then called HPOT) was evaluated and had positive feedback from the participants. A new website is in development for the product, now called the Tracking of Heavy Precipitation Objects (THePrO), and feedback will be collected on both the product and the website.

2.3 Description of Forecast Activities

A staple of FFaIR is the creation of a collaborative Day 1 ERO. The ERO is a probabilistic product issued by WPC for Days 1-3 operationally and Days 4-5 experimentally. The ERO highlights the categorical risk of rainfall exceeding flash flood guidance (FFG) within 25 miles of a point. In February 2022, the WPC updated the probabilities associated with the ERO, Table 3 shows what the old and new ERO risk probabilities are[3]. During the breakout sessions in the morning, one of the groups will be tasked with creating a collaborative ERO based on the new risk probabilities. An example of how the operational and experimental ERO can differ can be seen in Fig. 1 A-B.

Table 3: The old and new probabilities associated with WPC's ERO risk categories.

Risk	Old Probabilities	New Probabilities
Marginal (MRGL)	5-10%	5-15%
Slight (SLGT)	10-20%	15-40%
Moderate (MDT)	20-50%	40-70%
High (HIGH)	>50%	>70%

[3] For more information on this change and on the ERO please refer to the 2022 WPC Rainfall Overview.

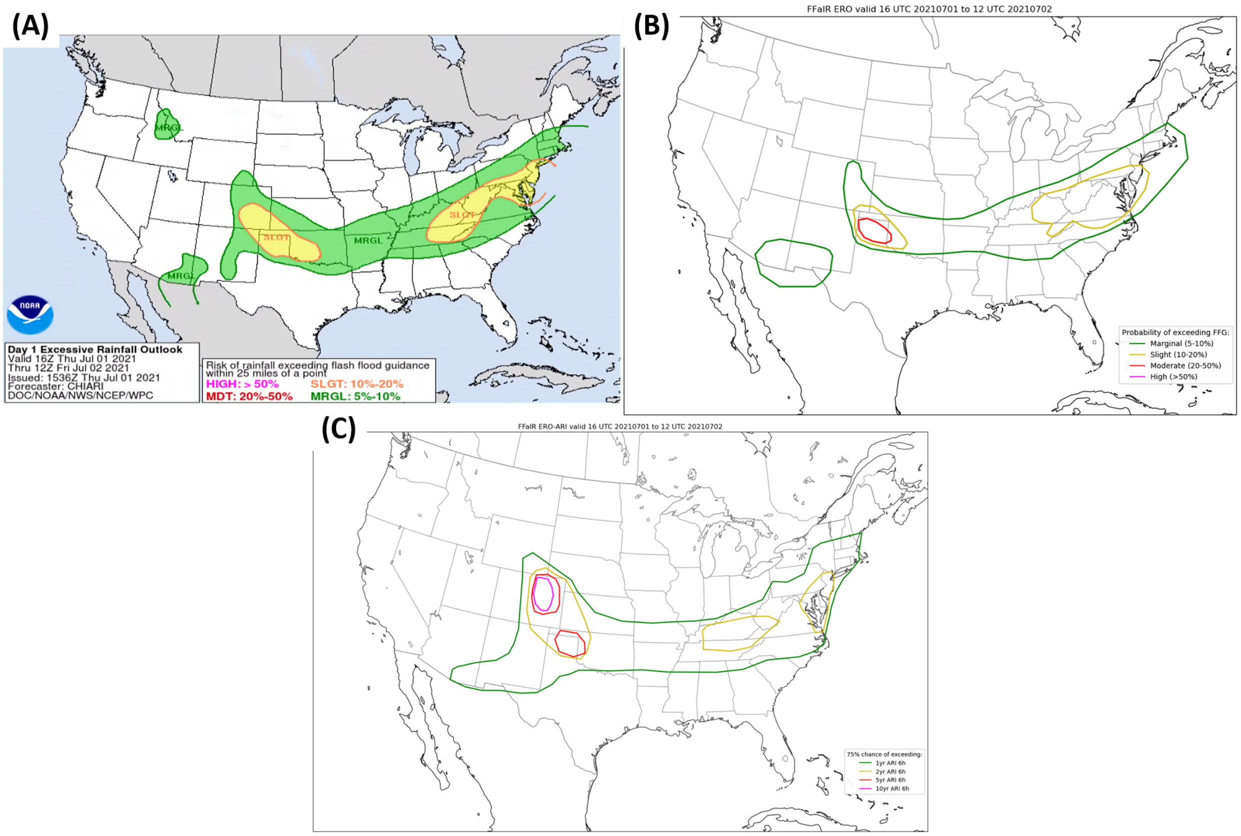

Figure 1: *The Day 1 (A) Operational ERO and (B) collaborative FFaIR ERO valid 16 UTC 01 July to 12 UTC 02 July, 2021. Marginal - green, slight - yellow, moderate - red, and high - pink. (C) The Day 1 ARI-ERO valid 16 UTC 01 July to 12 UTC 02 July, 2021. 75% chance of exceeding the 6 h:1y ARI - green, 2y ARI - yellow, 5y ARI - red, 10y ARI - pink.*

In addition to the creation of the ERO based on the traditional definition, last year FFaIR explored using ARIs to highlight areas at risk of excessive rainfall; this product was referred to as the ARI-ERO. Both products were valid from 16 UTC to 12 UTC. However, differing from the ERO, the ARI-ERO did not use risk categories. Instead, it highlighted where the six hour rainfall totals, sometime within the product's valid time period, had a 75% chance of exceeding the 1yr, 2yr, 5yr, and 10yr 6hr ARI. An example of this product can be seen in Fig 1C. Based on feedback from the participants and additional research on the utility of ARI as they relate to flood reports and extreme precipitation (ex. Lincoln and Thomason 2016, Herman and Schumacher 2018, Stovern et al. 2020, and Gourley and Vergara 2021), the ARI-ERO has been renamed and the definition altered for this year's FFaIR experiment. The renaming is due to feedback from last year that the name was too similar to the ERO and caused some confusion. Therefore, the ARI-ERO will now be referred to as AERO (pronounced arrow). Removing the "RI-" prevents the acronym for excessive rainfall outlook from standing out while still being present in the name of the product.

The ARIs that best correspond to extreme rainfall events and flash flooding varies depending on what QPE source is used and what datasets the ARI exceedances are verified against (i.e. just QPE or QPE and flood reports or QPE and flash flood warnings). Herman and Schumacher (2018) suggest that 24 hr QPE exceedances best relate to flash flood reports and warnings while Gourley and Vergara (2021) found better correspondence to flash flood reports at shorter accumulation times, 3 to 6 hrs. Similarly, Lincoln and Thomason (2018) found that 90% of flash flood reports could be captured using the 2 yr 3 hr ARI exceedance threshold. Since the AERO last year focused on 6 hr ARI exceedances and the majority of the research leans towards 3/6 hr ARI exceedances corresponding to flash flood reports, the AERO will once more be a product that highlights where forecasters identity the risk of various 6 hr ARIs being exceeded. The thresholds that will be used are: **2, 5, 10, 25 and 50 years.**

Additionally, the AERO, unlike last year, will not include a probability of exceeding the aforementioned thresholds in its definition (i.e. 75% probability of exceeding). This is due to participants seeming to focus too much on the probability value of the definition. Instead, the definition will be generalized to: "The most likely 6 hr ARI to be exceeded within 25 miles of a point." Lastly, although trivial, the colors used for the contours will be changed so that they do not match the colors used for the ERO risk categories. This change is needed since participants last year kept referring to the ARI thresholds as the ERO categories, which is not what the ARI thresholds are meant to be thought of as.

The other forecast activity will once again be the MRTP. At its core, the MRTP is a forecast for 6 hr rainfall totals and the identification of where the forecaster thinks the maximum rainfall will occur. However, the MRTP is designed to make the participants think about more aspects of the forecast than just precipitation amounts. This is done through activities like randomly assigning the participants a model or ensemble to evaluate while they are making, asking them what the impacts of the rainfall might be, or what ARI threshold they think will be exceeded. They are also tasked with completing a discussion on their forecast process and what they did/didn't like about the model/ensemble run they evaluated. The goal of such a product is to better understand what goes into a forecaster's process and to identify possible pitfalls or benefits of a model/ensemble system that are not easily seen in normal verification methods.

This year, the FFaIR team hopes to expand upon the MRTP by having the participants complete both a Day 2 and Day 1 forecast. There have been a few instances in the past two years in which participants were tasked to do this; an example of two MRTPs from last year can be seen in Fig. 2. However, these instances were random and "for fun", thus no real analysis was done on the Day 2 MRTPs when they were made. The vision for this exercise is to have the participants in the afternoon identify a region and 6 hour time period after 20 UTC the following day that has the threat for heavy rainfall and create a MRTP. The next morning, participants will use the updated guidance to create a MRTP for the region and valid time. Such a cadence will result in participants issuing two MRTPs each day, with the exception of the first day of the week. Such a process will help us evaluate the utility of the experimental guidance in the Day 2 time window.

Although this is the goal of the MRTPs, there are many factors that will impact whether or not issuing a Day 1 and Day 2 MRTP will be possible. The biggest factor will be how much of the experimental data will be provided out to or past forecast hour 60. An additional factor is whether we will have enough time each day to complete everything. It is very realistic that on some days, especially those with an active weather pattern or days the seminars are held, there simply will not be enough time in the day to complete two MRTPs. Additionally, as stated previously, virtual experiments tend to result in screen fatigue. Even if data and time allow for the issuance of two MRTPs, we do not wish to overwork our participants and if it seems as though screen fatigue is setting in, the Day 2 MRTP might be scratched.

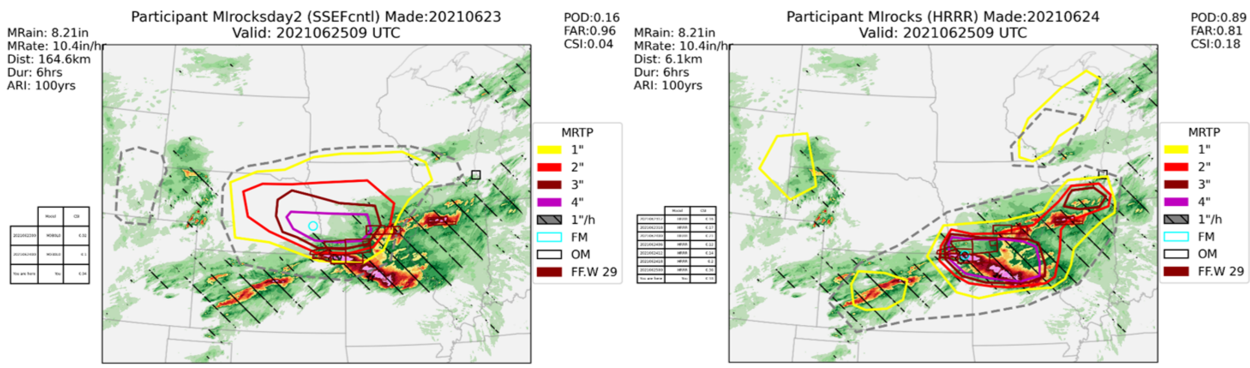

Figure 2: *Example of verification for a Day 2 (left) and Day 1 (right) MRTP valid 03 UTC to 09 UTC 25 June 2021, issued by MIrocks during the 2021 FFaIR experiment. Contours: yellow - 1 in., red - 2 in., dark red - 3 in., purple - 4 in., and dashed gray - 1in/h rainfall rates. The blue circle is the forecasted location of maximum rainfall. Black objects are the observed location with values listed in the upper left corner for: maximum rainfall (circle), 6 h ARI exceedance maximum (diamond), maximum in rainfall rate (square) and maximum duration of rainfall (star). Red polygons are Flash Flood Warnings issued during the event. On the left side of the images is a table listing the CSI for the forecaster, as well as for each run cycle that was valid during the event for the model/ensemble the forecaster was assigned. Top right of images: the forecaster's CSI, POD and FAR scores.*

3. Guidance, Products, and Data to be Evaluated

Over the past few years, the NWS has been working to transition all of its modeling systems to run on one dynamical core, the FV3[4] core. As development has progressed, the naming convention of the modeling systems has also changed, from being called the SAR[5] to the LAM[6] to the RRFS[7], the latter of which is the name of the system as a whole, not just the

[4] Finite Volume Cubed-Sphere
[5] Stand-Alone Regional
[6] Limited Area Model
[7] Rapid Refresh Forecast System

deterministic model. Additionally, the Environmental Modeling Center (EMC) and the Global Systems Laboratory (GSL) have been independently running their own versions of the RRFS. This is also true for the Storm-Scale Ensemble Forecast (SSEF) system that is run by the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma (OU). Although there are benefits to running multiple versions of FV3-CAMs that are independent of one another, there are also numerous drawbacks. Therefore, this year, including during HWT, the three data providers will be using the same control system as the base for their model/ensemble runs. The majority of the experiment will focus on the performance of these interdependent systems. In addition to these experimental model/ensemble runs, FFaIR will once more be evaluating the Machine-Learning Probabilities (MLP) Day 1 ERO products from Colorado State University (CSU). Also to be evaluated are a new version of WPC's Heavy Precipitation Object Tracker (HPOT). A general summary of the guidance and tools that will be evaluated in the 2022 FFaIR experiment can be seen in Table 4.

Table 4. *The deterministic and ensemble model guidance and products that will be evaluated in the 2022 FFaIR experiment (the experimental guidance is in the darker shade). Various acronyms can be found in the footnote #8*[8].

Provider	***Model***	***Resolution***	***Forecast Hours***	***Notes***
ESRL/GSL	***HRRR***	*3 km*	*Hourly forecasts. Forecast length: 00, 06, 12, and 18 UTC runs are 48 h. All other run times are 18 h, as is the sub-hourly output.*	*High resolution, hourly updated, convection allowing nest of the Rapid Refresh (RAP) model.*
EMC	***NAMnest***	*3 km*	*60 h forecast run daily at 00, 06, 12 and 18 UTC.*	*High resolution, convection allowing nest of the NAM.*

[8] 3DEnVar DA: 3 dimensional ensemble--variational data assimilation
EnKF: Ensemble Kalman Filter
RRFSDAS: Rapid Refresh Forecast System Data Assimilation System. It is a 30-40 member 3-km DA ensemble.
IC and LBC: Initial Conditions and Lateral Boundary Conditions
LSM: Land Surface Model

EMC	***HREFv3***	*Products output on 5km grid*	*48 h forecast run daily at 00, 06, 12 and 18 UTC.*	*Consists of 10 members, each member provides a real-time and time-lagged run.*
EMC/WPC/ MDL	***PQPF from GEFS, WPC, and NMB***	[redacted]	*Out to Day 3 at 00, 06, 12, and 18 UTC*	*6 h and 24 h QPF for the 90th, 95th, and 90th percentiles.*
EMC/GSL	***RRFSp1***	*~3 km*	*Hourly forecasts. 00 and 12 UTC runs out to 60 h.*	*Hourly analysis updates via hybrid 3DEnVar DA and partial cycling capability similar to that used by the NAM and its nests as well as RAP/HRRR.* ***RRFSp1 is the control system for RRFSp2.***
EMC/GSL	***RRFSp2***	*~3 km*	*00 UTC cycle only. Run out at least 36h, with possibility of longer runtime.*	*Hourly analysis updates via a hybrid 3DEnVar DA, using RRFSp1 as the initial state at 18z followed by six hours of cycling to 00z with the 3-km EnKF "RRFSDAS" ensemble.* ***RRFSp2 is the control member of the RRFSDAS.***
EMC/GSL	***RRFSe***	*~3 km*	*00 UTC cycle only. Run out at least 36h, with possibility of longer runtime.*	*A 9-member forecast. Initial conditions come from the experimental, real-time, hourly-cycled, 3-km EnKF "RRFSDAS" ensemble of RRFSp2.* ***RRFSe members 1-9 are perturbed forecasts initialized from the corresponding RRFSDAS members.***

CAPS	***CAPS_RRFSe***	*3 km*	*00 UTC run out to 84 h*	*10 to 20 member multi-physics and IC perturbation ensemble.* ***IC/LBC from the corresponding member of the RRFSDAS.***
CAPS	***RRFSp3*** *aka: M0B0L0_P*	*3 km*	*00 UTC run out to 84 h*	*No cycled DA.* ***Uses RRFSe mean for ICs.***
CAPS	***RRFSp4*** *aka: M0B0L0_PG*	*3 km*	*00 UTC run out to 84 h*	*No cycled DA. Same as RRFSp3 but ICs are pulled from the GEF*
CAPS	***RRFSp5 - RRFSp8***	*3 km*	*00 UTC run out to 84 h*	*No cycled DA.* ***Testing of various physics configurations.***

3.1 Experimental Models and Ensembles

As stated above, unlike in previous years, the experimental model and ensemble guidance provided by EMC, GSL, and CAPS will be interconnected to some extent. This is especially true for the two deterministic runs of the RRFS provided by EMC/GSL, which are referred to as RRFS prototypes 1 and 2; aka RRFSp1 and RRFSp2.

3.1.1 Deterministic

Table 5 provides a summary of the configurations for the deterministic models provided by EMC, GSL, and CAPS. The RRFSp1 features hourly analysis updates via a hybrid 3DEnVar data assimilation framework and has partial cycling similar to that used by the NAM and its nests as well as RAP/HRRR. The hybrid 3DEnVar algorithm leverages the freely available EnKF members from the Global Data Assimilation System (GDAS) to provide flow-dependent information in the EnVar cost function with no 3-km ensemble information used. The RRFSp1 will be run twice daily (00z and12z) out to forecast hour 60.

The RRFSp1 is the control system for RRFSp2 and is connected to RRFSp2 by providing high resolution, 3km central states at 18z to RRFSp2 around which the Global Ensemble Forecast System (GEFS) perturbations are re-centered. The input of initial state from the RRFSp1 is then followed by six hours of cycling to 00z with the 3-km EnKF "RRFSDAS[9]" ensemble providing flow-dependent information in the EnVar cost function during the hybrid analysis. The determinist forecast of the RRFSp2 is then launched at 00z. The RRFSp2 hybrid analysis is used to recenter the EnKF ensemble mean each hour thereby forming the control member of the EnKF ensemble. The hybrid analysis also includes RAP/HRRR-like analysis components including adjustments to the soil temperature and moisture along with a non-variational cloud and precipitation hydrometeor analysis. The current plan is to run this model out to at least 36 hours. After the completion of the HWT, there is a possibility that a 12z run will be added. If this occurs, the central states for the GEFS perturbations to re-center around will be provided by the RRFSp1 at 06z.

The 00z forecast of the RRFSp2 is then used as the ICs for the RRFSp3. This is considered the control member of the ensemble system that OU CAPS will be running. It, like the rest of the ensemble members, will be available out to 84 hours once daily (00z). In addition to the RRFSp3, OU CAPS is providing 5 additional deterministic configurations to help evaluate physics configuration; see Table 5. These will be referred to as RRFSp4-9. All the configurations have the same ICs/LBCs but the RRFSp4, which uses the GEFs. Additionally, RRFSp3 and RRFSp4 have the same physics configuration, so the only difference between the two models are their ICs and LBCs. Not all of these configurations will be evaluated in real time during FFaIR, but they will be included in our post-FFaIR analysis.

3.1.1.1 Changes After June 30 2022

Currently, EMC is running the RRFSp1 on their computing system. However, the center is in the middle of a transition from their old operational supercomputer to a new one. This transition is set to occur June 30. Unfortunately, the new system will likely not be ready to run the RRFSp1, which causes a hiccup in the cadence for running RRFSp2 and in turn in running RRFSp3. Therefore, GSL plans to stand up an identical version of RRFSp1 to run on their computing system; this will also be called RRFSp1. However, due to current computational resource limitations, at this time they are unsure how far out they will run the model, with it being extremely unlikely the forecasts will be out to 60 hours. The FFaIR team will adjust accordingly to what GSL decides and is able to do.

[9] Rapid Refresh Forecast System Data Assimilation System

Table 5: *The deterministic model configurations that will be evaluated in FFaIR 2022. *RRFSp1 evaluation might end June 30th, see section 3.1.1.1.*

Model	ICs	LBCs	Microphysics	PBL	LSM	Data Assimilation
RRFSp1	own	GFS – 00z and 12z cyc	Thompson-Eidhammer	MYNN	RUC	GFS cold start
RRFSp2	18Z RRFSp1 central state, hourly 3km hybrid 3DEnVar	GFS	Thompson-Eidhammer	MYNN	RUC	Hybrid 3DEnVar
RRFSp3 (M0B0L0_P)	RRFSe mean	GFS	Thompson	MYNN	NOAH	Inherited from the RRFSe
RRFSp4 (M0B0L0_PG)	GFS	GFS	Thompson	MYNN	NOAH	GFS cold start
RRFSp5 (M1B0L0_P)	RRFSe mean	GFS	NSSL	MYNN	NOAH	Inherited from the RRFSe
RRFSp6 (M0B0L1_P)	RRFSe mean	GFS	Thompson	MYNN	NOAH-MP	Inherited from the RRFSe
RRFSp7 (M1B2L2_P)	RRFSe mean	GFS	NSSL	TKE-EDMF	RUC	Inherited from the RRFSe
RRFSp8 (M0B2L1_P)	RRFSe mean	GFS	Thompson	TKE-EDMF	NOAH-MP	Inherited from the RRFSe

3.1.2 Ensembles

GSL will be providing a 10 member ensemble that will be referred to as RRFSe. Initial conditions come from the experimental, real-time, hourly-cycled, 3-km EnKF "RRFSDAS" ensemble of RRFSp2. The RRFSe members are the 9 first perturbed forecasts initialized from the corresponding RRFSDAS members along with the RRFSp2. The 9 members include stochastic parameter perturbations (SPP) applied to the land-surface, PBL, and microphysics schemes. A summary of the configuration for the RRFSe can be seen in Table 6. The ensemble is run once daily at 00z and has a forecast length out to at least 36 hours. As with the RRFSp2, there is a chance that after HWT is completed, GSL will begin running this at 12z as well.

The ensemble provided by OU CAPS, which will be referred to as CAPS_RRFSe, will be run once daily at 00z out to 84 hours. It will be an 11 member ensemble, similar to the one run for the 2022 Winter Weather Experiment. It will consist of the CAPs control member (RRFSp3). The other 10 members will get their ICs from the corresponding RRFSDAS member (i.e. CAPS_RRFSe member 1 has ICs from RRFSDAS member 1). Along with the variable ICs, the members will also have differing physics suites. The configuration for the ensemble can be seen in Table. 7.

Table 6: *The member configurations for the RRFSe system that will be evaluated in FFaIR 2022.*

Members	IC	LBC	Microphysics	PBL	LSM	Radiation
RRFSp2	See Table 5					
RRFSe01	enkf_m01	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe02	enkf_m02	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe03	enkf_m03	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe04	enkf_m04	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe05	enkf_m05	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe06	enkf_m06	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe07	enkf_m07	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe08	enkf_m08	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG
RRFSe09	enkf_m09	GEFS	Thompson-Eidhammer	MYNN	RUC	RRTMG

Table 7: *The member configurations for the CAPS_RRFSe system that will be evaluated in FFaIR 2022.*

Members	IC	LBC	Microphysics	PBL	Surface	LSM
M0B0L0_P (RRFSp3)	RRFSe mean	GFS	Thompson	MYNN	MYNN	NOAH
M0B0L0_PI	RRFS_01	GEFS_m1	Thompson	MYNN	MYNN	NOAH
M0B1L0_PI	RRFS_02	GEFS_m2	Thompson	Shin-Hong	GFS	NOAH
M0B2L1_PI	RRFS_03	GEFS_m3	Thompson	TKE-EDMF	GFS	NOAHMP
M0B0L1_PI	RRFS_04	GEFS_m4	Thompson	MYNN	MYNN	NOAHMP
M0B2L2_PI	RRFS_05	GEFS_m5	Thompson	TKE-EDMF	GFS	RUC
M1B0L0_PI	RRFS_06	GEFS_m6	NSSL	MYNN	MYNN	NOAH
M1B1L0_PI	RRFS_07	GEFS_m7	NSSL	Shin-Hong	GFS	NOAH
M1B2L1_PI	RRFS_08	GEFS_m8	NSSL	TKE-EDMF	GFS	NOAHMP
M1B0L1_PI	RRFS_09	GEFS_m9	NSSL	MYNN	MYNN	NOAHMP
M1B2L2_PI	RRFS_10	GEFS_m10	NSSL	TKE-EDMF	GFS	RUC

3.2 Machine Learning Products

3.2.1 CSU "First Guess" EROs

The machine learning group at CSU will once again be providing numerous "first guess" ERO MLPs. Training for these products is done using a random forest machine learning algorithm; for additional information on this please see Appendix A in [Herman and Schumacher 2018](). Like the WPC ERO, the MLP EROs are a probabilistic product. However, unlike the WPC ERO, the MLP EROs traditionally are collaborated to identify areas of excessive rainfall rather than the exceedance of FFG. Excessive rainfall is defined differently depending on the region of the CONUS[10] and is defined as either a flash flood report or a 2-year ARI exceedance or 1-year ARI exceedance occurs (Schumacher et al. 2021). The probabilities are based on a record of comparisons between historical model forecasts and precipitation observations.

Differing from last year, the MLP EROs will be centered around various versions of the GEFS trained ERO rather than different versions of the NSSL trained ERO. Currently, the version of the Day 1 GEFS MLP ERO (hereafter GEFSO) evaluated in the 2020 FFaIR experiment is operationally used at WPC. However, the GEFSO ERO was trained using the GEFSv11 (pre-FV3). When the GEFS transitioned from version 11 to version 12, the GEFSO was not retrained on the GEFSv12 reanalysis dataset as one was not available to be trained on when GEFSv12 was first implemented. Last year in FFaIR, it was found that despite this, the GEFSO Day 1 ERO still performed well. However, now that the GEFSv12 reanalysis is available, the CSU MLP group has retained their GEFS ERO; this version of the MLP will be referred to as the FV3GEFSR ERO. Aside from the using the updated reanalysis data for training, all other aspects of the ML training is identical to the GEFSO.

As discussed in Schumacher et al. 2021, for both the GEFSO and FV3GEFSR, the CONUS is divided into 8 different regions and the model is trained on those regions. Each region has a different definition of "excessive rainfall" used, which is a mix of flash flood and flood reports and Climatology-Calibrated Precipitation Analysis (CCPA) ARI exceedances; refer to Table 1 in Schumacher et al. 2021. The forecasts, made regionally, are then stitched together to create the final product. However, the MLP EROs are currently verified using a practically perfect method developed at WPC, which uses a set of observations referred to as the Unified Forecast Verification System (UFVS). The UFVS consists of Stage IV exceedances of FFG and the 5yr ARIs along with flash flood and flood LSRs and U.S. Geological Survey (USGS) river gauge observations. This led the CSU team to develop another GEFS-based ERO, referred to as the UFVSGEFSR. Like the FV3GEFSR, the UFVSGEFSR is trained using the GEFSv12 reanalysis data, but rather than using the original observational training dataset, it uses the UFVS

[10] The CONUS was divided into eight regions, with models trained separately for each of these regions.

dataset. It, like the other two versions, is trained on the 8 different regions but the observation dataset used for training is the same across all the regions.

All three GEFS ERO versions will be evaluated in FFaIR and the general differences among them can be found in Table 8. In the past, the CSU MLPs were only run once daily at 00z. However, at the request of WPC, CSU is now providing the FV3GEFSR and UFVSGEFSR at both 00z and 12z. The 12z runs arrive after 16 UTC, which is the start of the Day 1 ERO done during FFaIR. Therefore, the 12z EROs will not be used in real time. However, during the subjective verification portion of the experiment, the 00z and 12z forecast will be compared against one another.

Table 8: *CSU MLP EROs that will be evaluated during the 2022 FFaIR experiment.*

ERO MLP	Training Model	Forecast Model	Observational training set	Availability
GEFSO	GEFSv11	GEFSv12	Flood/Flash Flood LSRs and regional specific CCPA ARI exceedances	00z
FV3GEFSR	GEFSv12	GEFSv12	Flood/Flash Flood LSRs and regional specific CCPA ARI exceedances	00z and 12z
UFVSGEFSR	GEFSv12	GEFSv12	UFVS	00z and 12z
HRRR	HRRR	HRRR	Flood/Flash Flood LSRs and regional specific CCPA ARI exceedances	00z
NSSL2	NSSL_ARW	NSSL_ARW	Flood/Flash Flood LSRs and regional specific CCPA ARI exceedances	00z
BLEND	n/a	Weighed blend of the FV3GEFS, NSSL2 and HRRR	n/a	00z

In addition to the GEFS-based ERO, they will also be providing a Day 1 HRRR-based ERO as well as a Day 1 ERO that is a blend of the GEFSO, the NSSL-based ERO (NSSL2), and HRRR ERO (hereafter BLENDs). Both of these EROs were evaluated last year but it was found that the HRRR ERO lacked skill, especially across the southwest during the monsoon. It was hypothesized one reason for the model's difficulty in identifying the risk associated with the monsoon was due to the short training point (~2 years) and the fact that during the training period the SW Monsoon was weak to non-existent. Therefore, the model was retained using a larger dataset (~3.25 years) including last summer, which had an active monsoon. The 2022 version of the HRRR ERO also leverages the same spatial averaging of the predictors that is used in the NSSL2 ERO. This consists of averaging predictors spatially for each forecast gridpoint, rather than including values at selected surrounding grid points as separate predictors. The masking of grid points has also been updated. HRRRv2021 masked out points offshore during its training, while HRRRv2022 will do this as well as mask the grid points in Canada and Mexico.

The BLEND ERO is configured the same way as last year where each ML forecast (FV3GEFSR, NSSL2, and HRRR) is combined to create the ERO. The relative weight of each ML forecast is determined from their relative skill over the last 90 days. However, the GEFSO has been replaced with the FV3GEFS MLP and the HRRR ERO will be the updated version described above. Analyzing the performance of the NSSL2 is not a goal of the CSU team or FFaIR this year since the model has not been updated and it has been evaluated extensively since the 2020 FFaIR experiment; see pg 33 in the 2020 FFaIR Operations Plan for more information about NSSL2. However, since it is the baseline for CAM EROs and is included in the BLEND ERO, it will also be evaluated, acting as the baseline for the BLEND and HRRR to beat.

3.2.2 CAPS ML Ensemble Probabilities - HREF+

In addition to providing a RRFS-based ensemble, CAPS will also provide a new ML ensemble mean product for rainfall they have been working on. This product leverages a 12-member super-ensemble consisting of 4 FV3-LAM forecasts run by CAPS and 8 members of the HREF ensemble (4 members and their 12-hour time-lag counterparts). The product will be called the HREF+. The 4 members used from the HREF are the: hiresw_arw, hiresw_nssl, hrrr and NAMnest. The FV3 member was not used because its dataset to train on is too short. Due to constraints in the forecast hours covered by the HREF, these AI rainfall forecast products will be produced from 0 – 36 hours of forecast time. The forecasts will be produced daily for ensemble-based probability of rainfall exceeding 0.5 inches during a 6-h period between 0 and 36 hours of forecast time.

Individual ensemble member AI forecasts are generated using a U-Net (a deep learning approach which uses convolutional neural networks and is designed to identify spatial patterns in images). A set of 23 input variables are used from each forecast member, including predictions of wind, temperature, and moisture information at different vertical levels, as well as predictions of reflectivity, QPF, and precipitable water. The U-Net considers data over 64x64 patches, producing forecasts which are stitched together for each member to produce a CONUS forecast. A neighborhood maximum ensemble probability (NMEP) is applied to the collection of individual U-Net predictions for each member to produce the final product; the neighborhood is 45km and no gaussian smoother is used.

3.3 WPC Experimental Object Tracking Products

Two experimental products developed at the Development and Training Branch (DTB) at WPC will be evaluated. Both products utilize the Model Evaluation Tool (MET) Method for

Object-Based Diagnostic Evaluation (MODE) time-domain tool. The MODE-Time-Domain (MTD) tool applies an object-based technique (MET Online Tutorial). Using the MDT, the products identify 3-dimensional space/time objects and tracks the 2-dimensional objects through time. The Tracking of Heavy Precipitation Objects (THePrO), previously called the Heavy Precipitation Object Tracker (HPOT) has been evaluated in previous FFaIR experiments. The focus of evaluation for THePrO will be on a new website for hosting the product. The Analog Forecasting of Precipitation Objects (AnaPrO), formally the Heavy Precipitation Object Displacement (HPOD), was developed as a tool to be used alongside THePrO and has never been officially evaluated in FFaIR.

3.1.1 THePrO

As noted above, THePrO (aka HPOT) was evaluated in FFaIR in the past. This included evaluating the utility of the product and asking participants to provide feedback about the experimental website that was hosting it[11]. Based on feedback from the 2020 FFaIR experiments and from WPC forecasters, a new website was developed for the product. The main focus of the evaluation of the THePrO will actually be on the utility of the new site rather than on the product itself. However, a brief overview of the product is provided and the results from the previous FFaIR on the then named HPOT can be found in the 2020 FFaIR Final Report.

Understanding that forecasters often do not focus on a single pixel of heavy rainfall from a model simulation but rather they mentally assume a radius of possible impacts around the pixel, THePrO attempts to mimic this method by identifying heavy rainfall objects and track them through space and time to identify areas that might be impacted by heavy rainfall. The THePrO uses MTD to track these objects, which are defined as a region of moderate to heavy precipitation that covers an area of at least ~2000 km^2. It is a probabilistic tool that, once objects are identified in each member of the ensemble, simultaneously conveys information about QPF track location, size, timing, and intensity in one image. Figure. 3 shows an example of the product at f01 and f06. The image shows:

- The QPF object centroid and the model are identified by a marker symbol for some ensembles (right legend).
- The marker color indicates the 90th percentile of hourly accumulated precipitation within the object (left legend).
- The color scale on the right side of the product is the ensemble probability of being in a heavy precipitation object. When looking at the >= 0.1" per hour threshold, this blue shading represents the probability of >= 0.1" per hour for organized rainfall events.
- The lines extending from the object centroids indicate the projected track of the centroid.

[11] https://origin.wpc.ncep.noaa.gov/verification/mtd/view.php

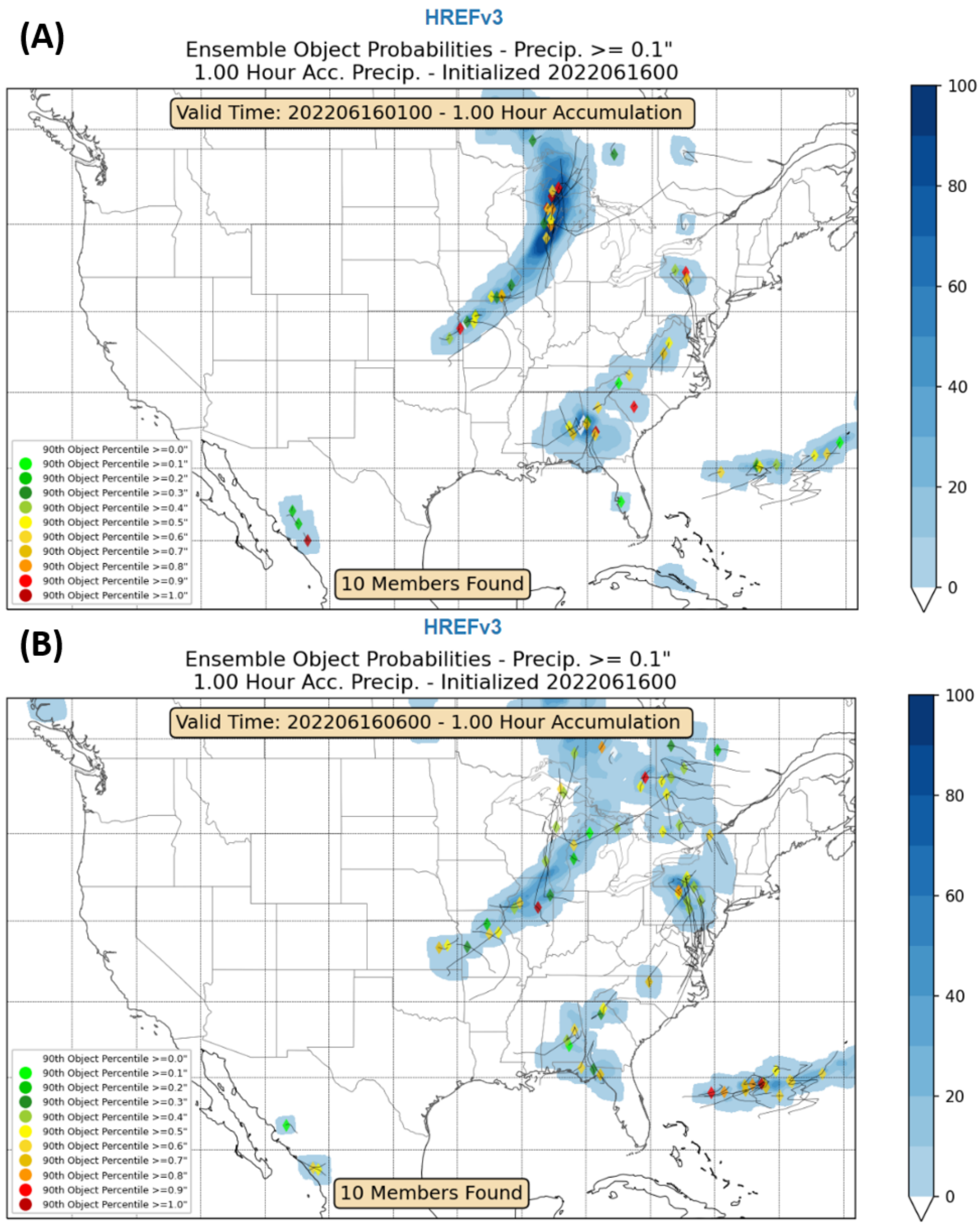

Figure 3: *The THePrO for the 00z run of the HREF valid (A) 01 UTC and (B) 06 UTC 16 June 2022. See above bullet points for description of legends and color bar.*

3.1.2 AnaPrO

The AnaPrO is an extension of the THePrO, expanding upon the product by providing displacement information for real-time objects identified by the THePrO and linking them to retrospective modeled objects and their observations. The displacement information is gathered for similar regions, intensities, time of year, and model forecast hours. Currently, this product is only available for the HRRR and it uses time-lagged members to create its ensemble. The AnaPrO currently provides three different probabilities: Full CONUS, By Object, and Centroid

Median Displacement. The latter of these will not be evaluated in FFaIR as it is still in its infancy. An example of the Full CONUS and By Object displacement probabilities can be seen in Fig. 4 and Fig. 5 respectively.

The Full CONUS probabilities conveys the probability of an object centroid falling within 40km of a point across the entire CONUS while taking into account the historical displacement biases. These probabilities are created by identifying similar retrospective objects to the current object. Similar objects are defined as objects within +/- 1 month, +/- 4 degrees of latitude and longitude, and +/- 12 hours lead time. Only objects that exist for 8 or more hours and can be linked to the retrospective archive are plotted in the product.

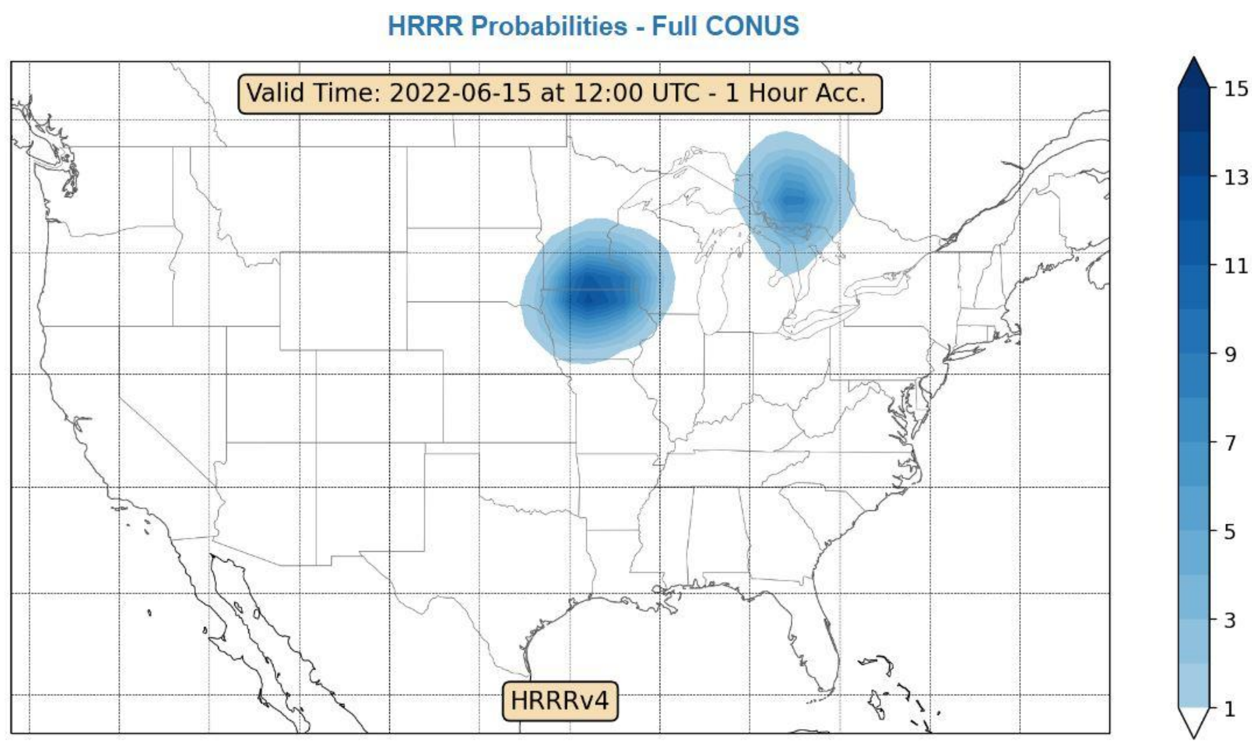

Figure 4: *The AnaPrO Full CONUS probabilities identifying the probability of object centroids occurring within 40km of a point.*

Using the objects identified by the Full CONUS analysis, the By Object plots focuses on each object to provide further information about displacement probabilities along with the probability that the object will exist in real time. It also includes object shape, centroid, and intensity details. For this example (Fig 5), the object that is being analyzed is the object along the MN/IA border in Fig 4. The image shows:

- The shape of the object, denoted by the blue shading. An important note is that **ALL** objects are plotted for this field, not just the one being analyzed.
- The object centroid path (black line) and marker color indicating the 90th percentile of hourly accumulated precipitation (left legend).

- Contour lines denoting the probability of object centroid location within 40 km of a point (also shown in "HRRR Probabilities - Full CONUS") given the retrospective object based verification. This verification is performed by comparing previously tracked model and observation objects.
    - For the image in Fig. 5, the object typically verifies southeast of the forecast centroid, but there is large spread, though the spread is relatively uniform.
- In the box to the northwest of the centroid, the probability that the observed object will exist (from the retrospective archive) for each model object. This is denoted by the "P = " and then the probability.

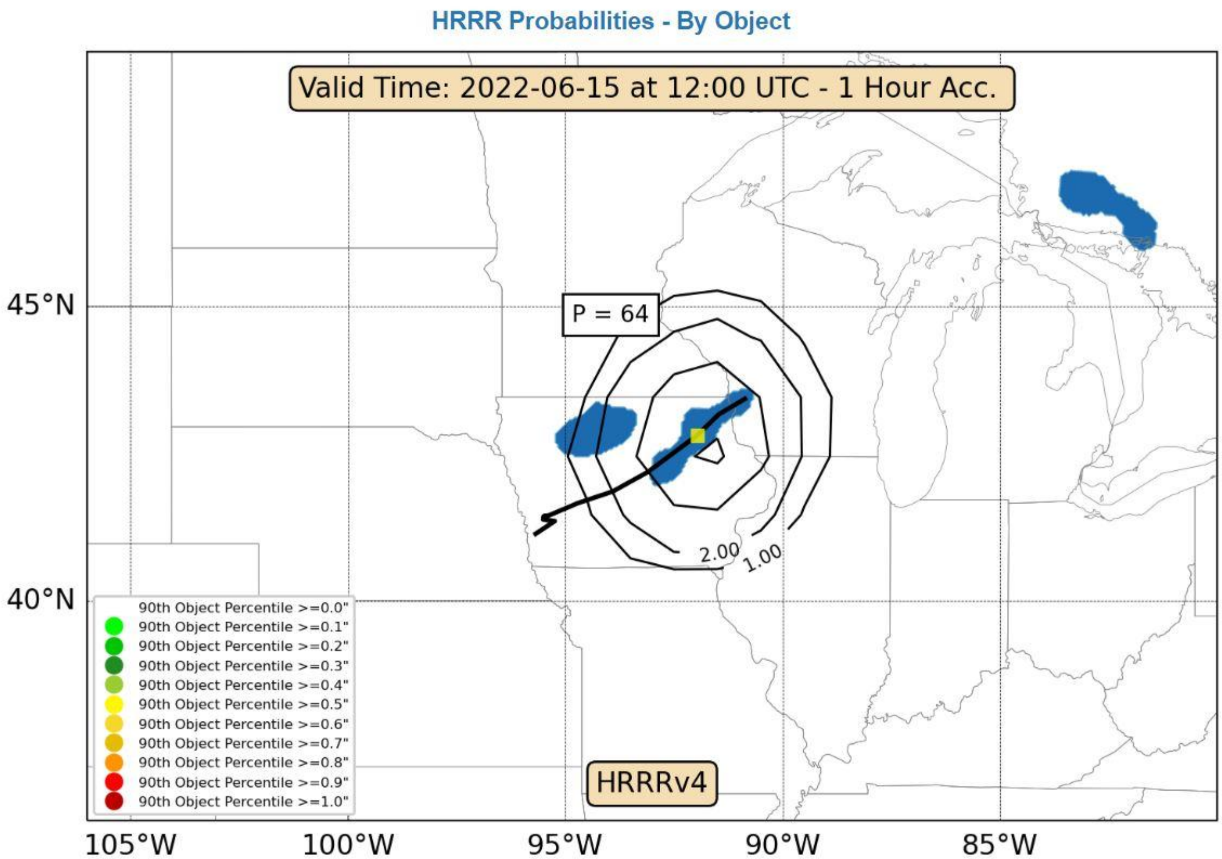

Figure 5: *The AnaPrO By Object Probabilities*

4. Summary

With the experiment once again being virtual, the FFaIR team will be relying heavily on the websites developed for the experiment to display the products being evaluated. This includes both a realtime and retrospective website and drawing tools to create the forecasts. Over the past two years these websites have slowly been upgraded and, as in the past, welcome feedback on how to make the sites more user friendly. Since this will be the third time in which FFaIR is completely virtual, the FFaIR team will use the lessons learned to reduce the mental exhaustion associated with day-long virtual meetings. This includes insisting that people step away from the screen when needed and reminding participants that they should treat this as if they were in-person for the experiment (i.e. put up an email out-of-office notice and don't push double duty by trying to perform daily work tasks).

Lastly, FFaIR once again will have an intern as part of the Bill Lapenta Internship. This year's intern is Alyssa Griffin. She is from Londonderry, New Hemisphere, and is currently attending Plymouth State University, majoring in meteorology. Alyssa will be helping with the daily duties of FFaIR, including running forecasting activities. Her research project will be an extension of last year's intern Miranda's work on modeling precipitation rates. Since the experimental models are now also outputting the maximum hourly precipitation rate, she will be focusing on the maximum rates rather than just the instantaneous rates.